

Behavior of vegetation sampling methods in the presence of spatial autocorrelation

Sarah C. Goslee

USDA-ARS Pasture Systems and Watershed Management Research Unit, Bldg. 3702 Curtin Road, University Park, PA 16802, USA; (e-mail: Sarah.Goslee@ars.usda.gov; phone: +814-863-0887; fax: +814-863-0935)

Received 5 March 2004; accepted in revised form 25 February 2005

Key words: Cover, Frequency, Modified Whittaker plot, North Carolina Vegetation Survey, Species-area curves, Transect

Abstract

Spatial autocorrelation in vegetation has been discussed extensively, but little is yet known about how standard plant sampling methods perform when confronted with varying levels of patchiness. Simulated species maps with a range of total abundance and spatial autocorrelation (patchiness) were sampled using four methods: strip transect, randomly located quadrats, the non-nested multiscale modified Whittaker plot and the nested multiscale North Carolina Vegetation Survey (NCVS) plot. Cover and frequency estimates varied widely within and between methods, especially in the presence of high patchiness and for species with moderate abundances. Transect sampling showed the highest variability, returning estimates of 19–94% cover for a species with an actual cover of 50%. Transect and random methods were likely to miss rare species entirely unless large numbers of quadrats were sampled. NCVS plots produced the most accurate cover estimates because they sampled the largest area. Total species richness calculated using semilog species-area curves was overestimated by transect and random sampling. Both multiscale methods, the modified Whittaker and the NCVS plots, overestimated species richness when patchiness was low, and underestimated it when patchiness was high. There was no clear distinction between the nested NCVS or the non-nested modified Whittaker plot for any of the measures assessed. For all sampling methods, cover and especially frequency estimates were highly variable, and depended on both the level of autocorrelation and the sampling method used. The spatial structure of the vegetation must be considered when choosing field sampling protocols or comparing results between studies that used different methods.

Introduction

Early ecologists assumed that vegetation was uniform, or oriented along a simple gradient, and designed their sampling methods accordingly (Greig-Smith 1979; Legendre and Fortin 1989). Methods employing a series of equally-sized quadrats, whether placed systematically or

randomly located, were widely used. The major concerns were with the number, size and shape of quadrats needed to characterize a particular vegetation type (Clapham 1932; Cain 1938; Rice and Kelting 1955). Ecologists now realize that most if not all communities are spatially structured, and that sampling and analysis methods must be robust to varying levels of spatial autocorrelation as

expressed by patchiness in the underlying species distributions (Levin 1992; Legendre 1993).

Most practitioners now realize that spatial autocorrelation interferes with standard statistical tests, and are familiar with statistical methods to overcome this complication (e.g. Dale and Fortin 2002; Legendre et al. 2002). Considerable effort has been devoted to developing sampling designs and statistical techniques to detect and quantify spatial pattern, especially as it relates to the measurement of species diversity (Fortin et al. 1989; Dutilleul 1993; Bellehumeur and Legendre 1998; Legendre et al. 2002). Although the potential importance of spatial autocorrelation in field sampling has been acknowledged (Stohlgren et al. 1995), little quantitative information is available about the effect of patchiness in vegetation on the measurement of common community attributes.

Ecologists continue to debate the appropriate shape of sampling areas, but attention has shifted to sampling schemes that incorporate subplots of a range of sizes. Multiscale methods are better-suited to patchy environments, and reduce or eliminate concerns about choosing the appropriate spatial resolution (Peet et al. 1998). The increased use of multiscale methods reflects interest in patterns of species richness and abundance at different scales (Stohlgren et al. 1995; Peet et al. 1998). Both nested plot designs, with overlapping subquadrats, and non-nested plot designs have been proposed. The modified Whittaker plot, developed by Stohlgren et al. (1995), is a widely-adopted non-nested method. All subplots are contained within the outermost sample region, but there is no overlap between other subplots, so I have described this as 'non-nested' even though it is truly minimally nested (Figure 1a). The NCVS uses a nested multiscale design with complete overlap between subplots of different sizes (Figure 1b; Peet et al. 1998). These newer methodologies have been developed to match advances in our understanding of the nature of spatial structure in vegetation, but both systematic and random quadrat methods are still widely used, especially when little is known in advance about the structure of the vegetation (Legendre et al. 2002).

Although multiscale designs are intended to address spatial patterning in vegetation, there has been little systematic comparison of their benefits in the presence of varying levels of autocorrelation. There have been a number of studies com-

paring aspects of the performance of vegetation sampling methods under field conditions, but none have explicitly examined autocorrelation (e.g. Bourdeau 1953; Stohlgren et al. 1998; Barnett and Stohlgren 2003; Korb et al. 2003; Leis et al. 2003). Instead of repeating these trials with field data, I chose to use simulated species distributions. Unlike field sampling, with simulated data the level of autocorrelation, the abundance of each species and the actual number of species are known exactly, so the performance of each sampling design under a range of conditions can be assessed.

The main objective of this study was to compare species cover and species frequency estimated by four methods, two quadrat (random and systematic) and two multiscale (nested and non-nested), given plant species of different size, abundance, and level of spatial autocorrelation. A secondary objective was to examine the effects of autocorrelation on species-area curves and estimates of total richness. This information provides a baseline for comparison of results from studies using different methods, and for evaluating the suitability of a particular sampling design for a variety of conditions.

Methods

Four sampling methods were selected: systematically-located quadrats, randomly-located quadrats, a non-nested multiscale and a nested multiscale scheme. The systematic quadrat method used a strip transect of 1×1 m quadrats running the length of the simulated area, for a total of fifty 1 m^2 quadrats. The start point of the transect was randomly chosen. The random method used the same size and number of quadrats, but they were placed randomly with no overlap. Species cover was recorded for each quadrat.

The non-nested design used was the modified Whittaker plot (Figure 1a; Stohlgren et al. 1995). The total area sampled was 20×50 m. Species cover was recorded within ten 2×0.5 m quadrats. Species presence was recorded for progressively larger rectangular plots: two of 2×5 m, one of 5×20 m, and a final plot of 20×50 m that encompassed all the smaller plots.

The NCVS design was a nested design that covered the same total area as the modified Whittaker plot, but divided the area differently

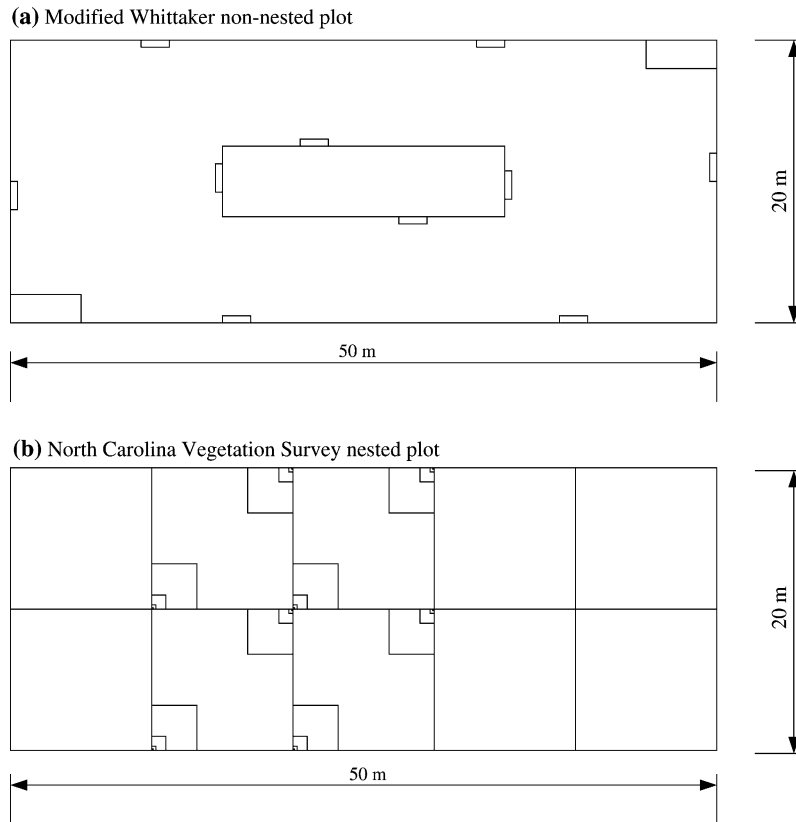


Figure 1. Multiscale sampling methods. (a) Modified Whittaker plot sampling method (Stohlgren et al. 1995). Cover is estimated within ten 1 m^2 quadrats, and species lists are recorded for each of the larger plots. (b) North Carolina Vegetation Survey (Peet et al. 1998) scheme. Cover is recorded within ten 100 m^2 plots, and species lists are recorded for each of the smaller subplots.

(Figure 1b; Peet et al. 1998). The entire $50 \times 20 \text{ m}$ area was divided into ten $10 \times 10 \text{ m}$ plots, in which cover was recorded. Four of these plots contained two sets of square nested subplots with areas of 10, 1, 0.1, and 0.01 m^2 . Species presence was recorded for these subplots.

Simulations

The *gstat* package for R was used to simulate species distributions with varying spatial autocorrelations, revealed as patchiness in the distribution of individuals (Pebesma 2001, 2004; R 1.8, R Project for Statistical Computing). The spatial correlation was quantified by an exponential variogram model, and ten levels of patchiness were created by varying the effective range (range = 0, 2, 5, 10, 15, 20, 25, 35, 50, 100). The variogram was used as the basis for an unconditional kriging

simulation at gridded prediction locations. The sequential gaussian simulation algorithm in the *gstat* package was used to generate multiple realizations at the grid notes. The resulting continuous variables were sliced to create simulated species with one of eight possible levels of abundance by defining all grid cells less than the cutoff value to be occupied, and the remainder to be vacant (abundance = 0.1, 1, 5, 10, 25, 50, 75, 90% cover). For binary samples such as these derived presence/absence grids, the level of spatial autocorrelation or patchiness is shown by ‘clumping’ of the species distributions across the simulated area (Figure 2).

The vegetation sampling schemes consisted of plots and subplots of predetermined sizes, so a scale had to be imposed on the simulated data. A grid cell or pixel is the lower limit of resolution of the sampling (the minimum size of a plant). Preliminary simulations using a range of resolutions (grain sizes of 1–20 pixels/m), conducted by sim-

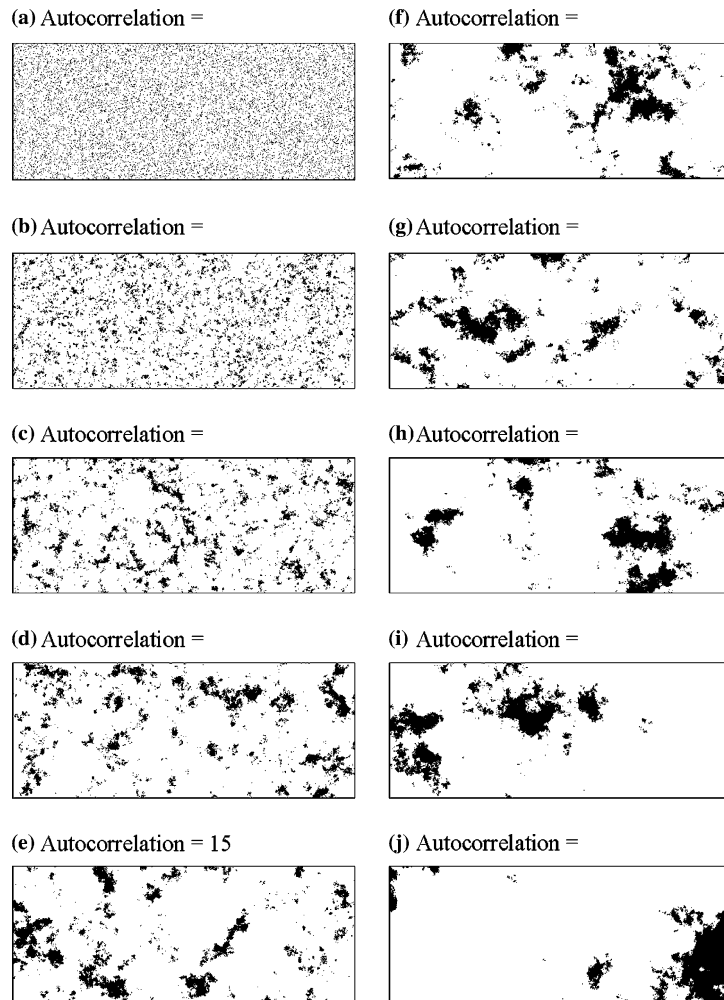


Figure 2. Simulated species distributions with 10% abundance and ten levels of autocorrelation (patchiness) determined by the range value for an exponential variogram.

ulating a 20×50 m grid of the appropriate number of pixels, demonstrated that varying the plant size within this range made little difference to the patterning of the simulated vegetation, so only the results of the 10 pixel/m simulations are presented here (plants with a minimum size of 10×10 cm). The grid used was thus 200×500 pixels, since the maximum size of a sampling plot was 20×50 m.

The single-species simulations used 10 levels of patchiness and 8 species abundances (80 combinations), and each had 100 repetitions. The four sampling methods were applied to each, so that total frequency and total cover as estimated by each could be compared to the actual values.

Simulated species were combined into 'communities' by randomly selecting two species from each abundance level for a given level of autocorrelation. This process resulted in fifty sets with 16 species of varying abundance but identical levels of patchiness. For the random and transect data, the cumulative species richness curve was used to estimate total richness in the 20×50 m area. Species-area curves were calculated from the subplots of the modified Whittaker and NCVS schemes, and again used to estimate total species richness. Both semilog [$richness \sim \log(area)$] and log-log [$\log(richness) \sim \log(area)$] forms of the species area curve were calculated.

Results and discussion

The NCVS plot provided a perfect mean cover estimate because cover is measured across the entire area (mean of ten 100 m² quadrats). The transect and random sample estimates are the means of fifty 1 m² quadrats, and the modified Whittaker estimate is the mean of ten 1 m² quadrats. If no spatial autocorrelation is present, these three methods give equivalent values with high accuracy and precision, as shown in Figure 3a. As the level of patchiness increases, the range of the species cover estimates also increases. Transect sampling was most affected by autocorrelation; species with a high autocorrelation and true abundance of 50% had measured cover of 19–94% from fifty quadrats (Figure 3b). Random sampling was less affected, with a range of 36–64%. Because it only samples ten 1 m²

quadrats, the modified Whittaker plot produced less-accurate cover estimates than the random sample, although it did perform better than the transect sample (range 28–76%). When only ten quadrats from the random and transect samples were considered instead of all fifty, the modified Whittaker plot and the random samples provided similar estimates of cover, and both were much more accurate than the transect sample.

Frequency estimates were more affected by the level of patchiness than by sampling method. Estimates from transect sampling had the greatest variability, followed by random and modified Whittaker, and NCVS estimates were the most precise (Figure 4). Frequency estimates from NCVS sampling tended to be higher than those obtained by other methods because of the larger size of the subplots.

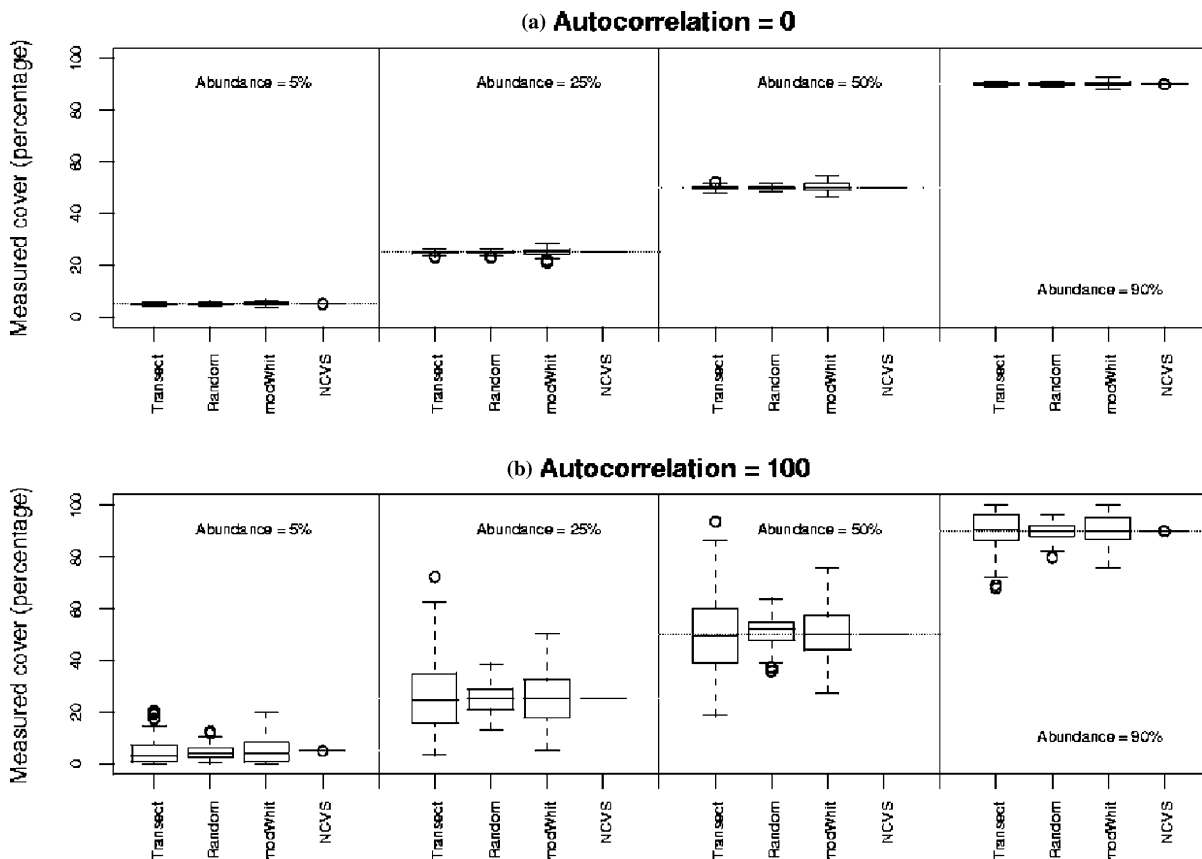


Figure 3. Comparison of strip transect, random quadrat, modified Whittaker plot and North Carolina Vegetation Survey sampling methods at low (range = 0) and high (range = 100) levels of autocorrelation for four levels of species abundance.

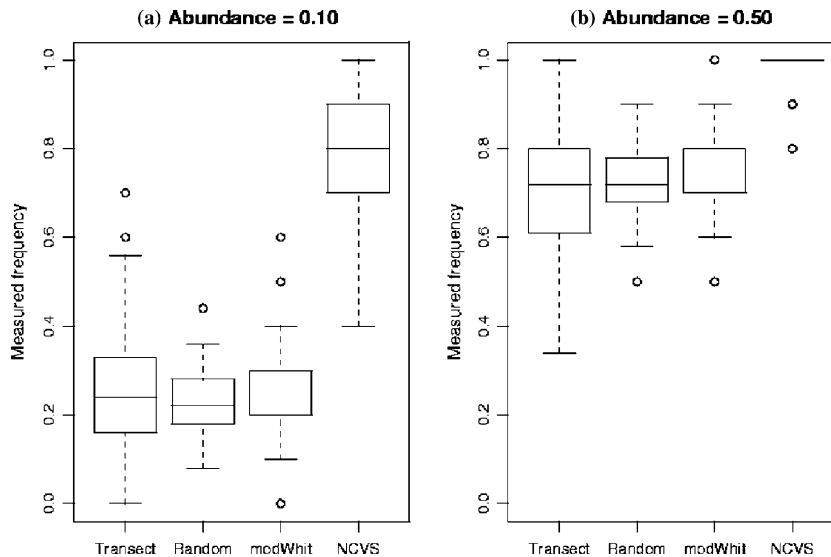


Figure 4. The range of measured frequencies obtained by using strip transect, random quadrat, modified Whittaker plot and North Carolina Vegetation Survey sampling methods on simulated species with high autocorrelation level and two levels of abundance.

Rather than present similar data for all four methods, I chose random sampling as representative of the effects of autocorrelation on measured plant frequency (Figure 5). The more uniform the distribution, the higher the frequency. With no autocorrelation, all species with abundance $> 10\%$ had a measured frequency of one using any method. At very high abundances, the frequency was near one regardless of the autocorrelation, while the frequency estimates for low and moderately abundant species varied widely. The effects of spatial autocorrelation on apparent cover and frequency were most pronounced at intermediate abundances, since 0 and 100% provide hard limits on plant cover. Rare species are unlikely to be found regardless of the sampling method or spatial distribution, while abundant species will always be found.

The format of the modified Whittaker and NCVS plots is fixed, but researchers often choose a certain number of quadrats for random and transect sampling. Simulation results can help to choose an appropriate number of quadrats (Table 1). Random sampling is slightly more efficient than transect sampling, but both are poor at locating rare species, especially if any autocorrelation is present. With random sampling from a species distribution based on a variogram with range = 20 and an abundance of 1%, at least 30

quadrats are required to have a 95% chance of encountering that species, and with transect sampling at least 47 quadrats are required. Both methods will find common species, even with few quadrats. The numbers of quadrats needed to find rare species are higher than often used to sample an area of 20×50 m, suggesting that these methods seriously underestimate species richness. Field studies support this conclusion; Jorgensen and Tunnell (2001) found that quadrat methods missed up to 30% of the plant species in their test area.

The multiscale plots can be used to estimate frequency at a range of spatial scales (Figure 6). Frequencies are always lower in smaller subplots, and patchiness has the greatest effect on the smallest subplots. The only exception is the 0.01 m^2 subplot of the NCVS scheme. This is the same size as the pixel used in the simulations, so frequency at this scale approximates the species abundance. Similarly-sized subplots follow the same curves for both modified Whittaker and NCVS sampling schemes.

Log-log species area curves substantially underestimated the number of species in 1000 m^2 , so only semi-log curves will be considered further. Cumulative richness curves from both transect and random plots overestimated the number of species in a 1000 m^2 area, especially at moderate and high levels of autocorrelation (Figure 7). This was

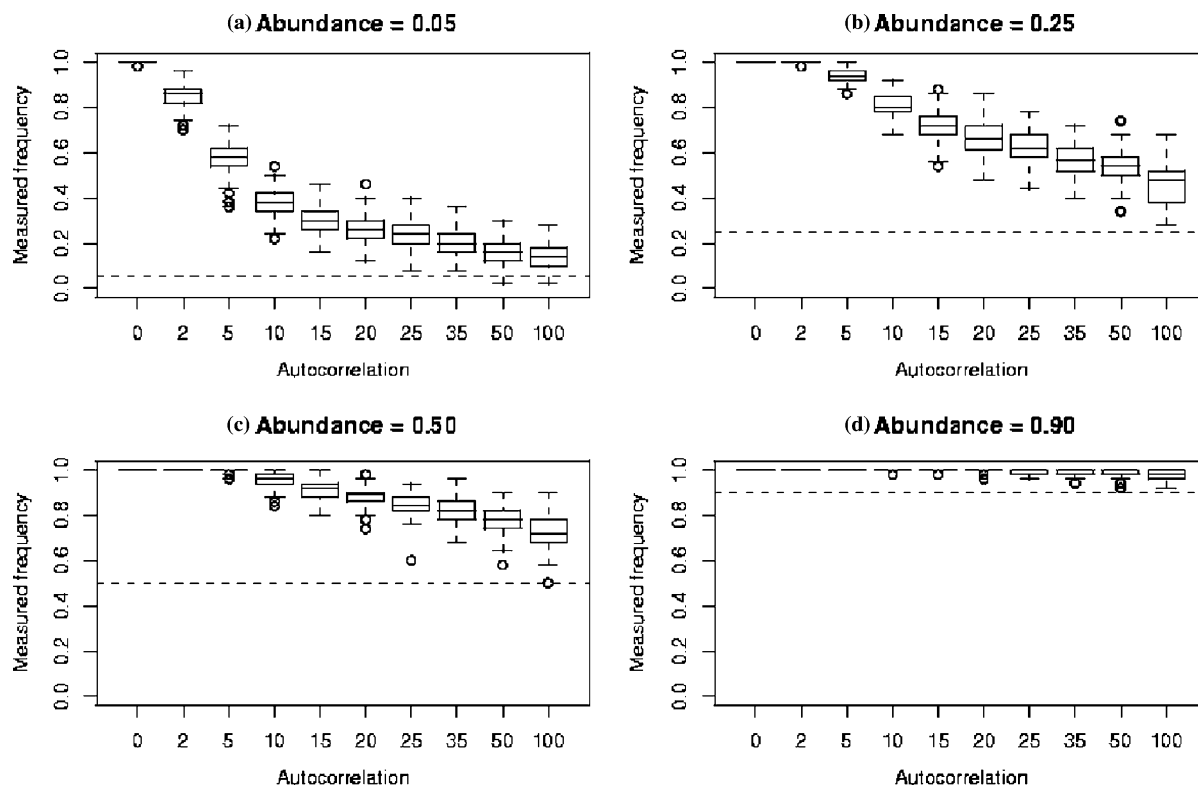


Figure 5. The range of frequency estimates produced by random sampling of simulated species with four abundance levels at a range of autocorrelation levels.

expected, since it is known that species-area curves constructed by aggregating samples have a higher slope, and thus overestimate total richness (Rosenzweig 1995). Although the shape of the response was somewhat different, both of the multiscale sampling schemes overestimated the total species richness at low autocorrelations, but underestimated it at high levels. For both, a uniform species distribution led to a species-area curve with low slope but high intercept, since even the smallest plots found most species. With high autocorrelation the slope was steeper but the intercept was much lower.

The 'communities' used for calculating species area curves are not representative of actual communities. Usually there are a few abundant species and many rare species, rather than equal numbers across a range of abundances. Each species was randomly distributed with respect to all other simulated species, and all had the same level of autocorrelation. In actual plant communities, species distributions may show positive and neg-

ative associations and the effects of environmental heterogeneity. The objective of this study was to evaluate the effects of autocorrelation, not hypotheses about the structure of plant communities, so this simplified community structure was appropriate.

The modified Whittaker plot was the only method examined with rectangular rather than square quadrats. Preliminary investigation showed that for this simulation study, plot shape made little difference. In the presence of anisotropy, such as an underlying environmental gradient, plot shape could have a strong effect on observed vegetation characteristics.

The choice of nested or non-nested plots has been debated in the literature. The only differences between the nested NCVS and the non-nested modified Whittaker plots that were not attributable to differing subplot sizes was in the species-area curve predictions of total richness. Stohlgren et al. (1995) stated that spatial autocorrelation has a larger effect on overlapping subplots because a

Table 1. The number of quadrats (maximum of 50) needed to have at least a 95% chance of encountering a species with given abundance and distribution.

Autocorrelation	Species abundance (%)							
	0.1	1	5	10	25	50	75	90
a. Transect								
0	27	3	1	1	1	1	1	1
2	46	5	2	1	1	1	1	1
5	—	14	5	4	1	1	1	1
10	—	28	9	6	3	2	1	1
15	—	48	16	12	4	1	1	1
20	—	47	24	14	9	4	1	1
25	—	—	32	15	6	3	1	1
35	—	—	38	23	10	3	2	1
50	—	—	38	30	14	6	4	1
100	—	—	—	42	27	13	4	1
b. Random								
0	23	3	1	1	1	1	1	1
2	—	5	2	1	1	1	1	1
5	—	13	4	3	2	1	1	1
10	—	16	6	4	2	1	1	1
15	—	29	10	5	3	2	1	1
20	—	30	10	7	3	2	1	1
25	—	39	15	7	3	2	1	1
35	—	48	13	10	4	3	1	1
50	—	—	19	10	5	2	1	1
100	—	—	22	11	6	3	2	1

A dash marks denote species which were not found after sampling 50 quadrats.

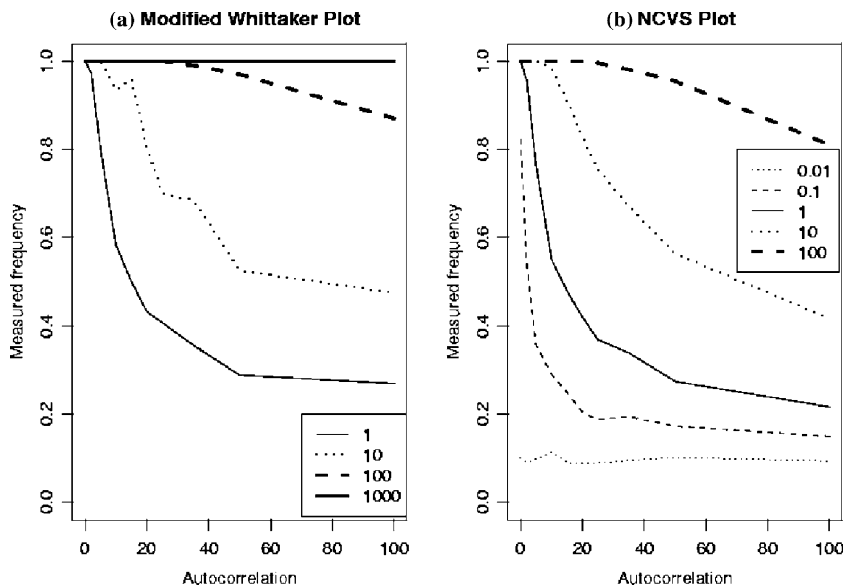


Figure 6. Frequency estimates produced by sampling at different spatial scales using the modified Whittaker and North Carolina Vegetation Survey schemes for species with 10% actual abundance.

particular area is sampled more than once. Peet et al. (1998) made the identical claim in reverse, stating that the use of non-overlapping plots

ignores spatial autocorrelation in vegetation. The simulation results here suggest that the two methods do respond differently to spatial

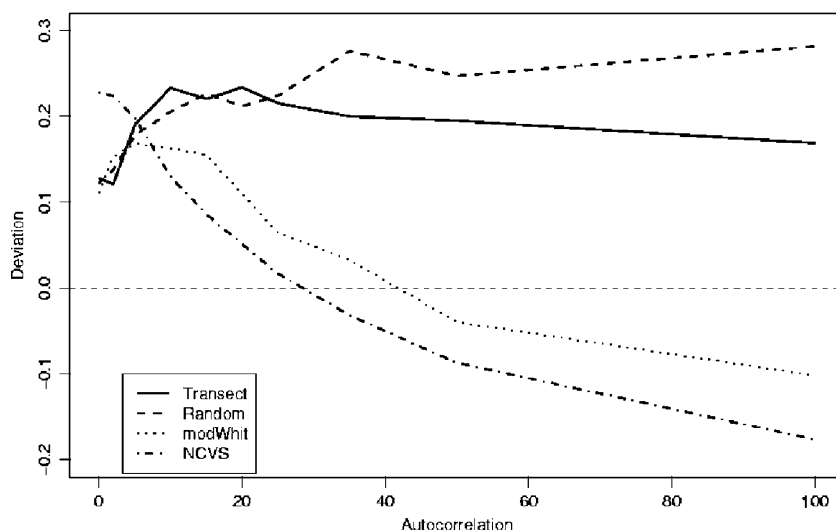


Figure 7. Deviances [(observed value–expected value)/expected] for species richness in 1000 m² estimated from data collected using four different sampling methods.

autocorrelation, but that the differences are negligible in anisotropic environments. It is more important to note that both methods overestimate species richness in vegetation with low autocorrelation, but underestimate it with high autocorrelation.

The findings from this simulation study on cover and frequency estimates accord with what is known from field experiments. All sampling methods encounter dominant species, but random and transect methods are more likely to miss rare species than methods that cover a larger area (Bourdeau 1953; Stohlgren et al. 1998; Jorgensen and Tunnell 2001; Barnett and Stohlgren 2003; Korb et al. 2003). The drawback is that larger multiscale methods are slower to lay out and sample in the field (Stohlgren et al. 1998; Korb et al. 2003).

Conclusions

In vegetation, the degree of underlying spatial autocorrelation is expressed by patchiness in the species distribution, and can be modeled by varying the effective range of the variogram used as the basis for a kriging simulation. Using simulated species distributions to test the choice of sampling method demonstrated that transect methods especially responded poorly to spatial autocorrelation. Randomly-located quadrats were more

efficient in highly patchy environments, but large numbers of quadrats are needed for either method if rare species must be located. Multiscale methods were more robust to spatial autocorrelation, and more effective at identifying rare species because of the larger proportion of the total area sampled. Cover and especially frequency estimates were highly variable, and depended on both the level of autocorrelation and the sampling method used. It is difficult to meaningfully compare community attributes obtained by different methods, and under certain circumstances sampled results may have little resemblance to actual values.

References

- Barnett D.T. and Stohlgren T.J. 2003. A nested-intensity design for surveying plant diversity. *Biodivers. Conserv.* 12: 255–278.
- Bellehumeur C. and Legendre P. 1998. Multiscale sources of variation in ecological variables: modeling spatial dispersion, elaborating sampling designs. *Landscape Ecol.* 13: 15–25.
- Bourdeau P.F. 1953. A test of random versus systematic ecological sampling. *Ecology* 34: 499–512.
- Cain S.A. 1938. The species-area curve. *Am. Midl. Nat.* 19: 573–581.
- Clapham A.R. 1932. The form of the observational unit in quantitative ecology. *J. Ecol.* 20: 192–197.
- Dale M.R.T. and Fortin M.-J. 2002. Spatial autocorrelation and statistical tests in ecology. *Écoscience* 9: 162–167.
- Dutilleul P. 1993. Spatial heterogeneity and the design of ecological field experiments. *Ecology* 74: 1646–1658.

- Korb J.E., Covington W.W. and Ful P.Z. 2003. Sampling techniques influence understory plant trajectories after restoration: An example from Ponderosa pine restoration. *Restor. Ecol.* 11: 504–515.
- Fortin M.-J., Drapeau P. and Legendre P. 1989. Spatial autocorrelation and sampling design in plant ecology. *Vegetatio* 83: 209–222.
- Grieg-Smith P. 1979. Pattern in vegetation. *J. Ecology* 67: 755–779.
- Jorgensen E.E. and Tunnell S.J. 2001. The effectiveness of quadrats for measuring vascular plant diversity. *Tex. J. Sci.* 53: 365–368.
- Legendre P. 1993. Spatial autocorrelation: trouble or new paradigm? *Ecology* 74: 1659–1673.
- Legendre P. and Fortin M.-J. 1989. Spatial pattern and ecological analysis. *Vegetatio* 80: 107–138.
- Legendre P., Dale M.R.T., Fortin M.-J., Gurevitch J., Hohn M. and Myers D. 2002. The consequences of spatial structure for the design and analysis of ecological field surveys. *Ecography* 25: 601–615.
- Leis S.A., Engle D.M., Leslie D.M.Jr., Fehmi J.S. and Kretzer J. 2003. Comparison of vegetation sampling procedures in a disturbed mixed-grass prairie. *Proc. Oklahoma Acad. Sci.* 83: 7–15.
- Levin S.A. 1992. The problem of pattern and scale in ecology. *Ecology* 73: 1943–1967.
- Pebesma E.J. 2001. *GSTAT User's Manual*. (gstat version 2.3.3). <http://www.gstat.org>.
- Pebesma E.J. 2004. Multivariable geostatistics in S: the gstat package. *Comput. Geosci.* 30: 683–691.
- Peet R.K., Wentworth T.R. and White P.S. 1998. A flexible, multipurpose method for recording vegetation composition and structure. *Castanea* 63: 262–274.
- Rice E.L. and Kelting R.W. 1955. The species-area curve. *Ecology* 36: 7–11.
- Rosenzweig M.L. 1995. *Species Diversity In Space and Time*. Cambridge University Press, Cambridge, UK.
- Stohlgren T.J., Falkner M.B. and Schell L.D. 1995. A Modified-Whittaker nested vegetation sampling method. *Vegetatio* 117: 113–121.
- Stohlgren T.J., Bull K.A. and Otsuki Y. 1998. Comparison of rangeland vegetation sampling techniques in the Central Grasslands. *J. Range Manage.* 51: 164–172.